YABLO'S PARADOX AND ARITHMETICAL INCOMPLETENESS

GRAHAM LEACH-KROUSE

Abstract. In this short paper, I present a few theorems on sentences of arithmetic which are related to Yablo's Paradox[1] as Gödel's first undecidable sentence was related to the Liar paradox.[2] In particular, I consider two different arithemetizations of Yablo's sentences: one resembling Gödel's arithmetization of the Liar, with the negation outside of the provability predicate, one resembling Jeroslow's undecidable sentence,[3] with negation inside. Both kinds of arithmetized Yablo sentence are undecidable, and connected to the consistency sentence for the ambient formal system in roughly the same manner as Gödel and Jeroslow's sentences.

Finally, I consider a sentence which is related to the Henkin sentence "I am provable" in the same way that first two arithmetizations are related to Gödel and Jeroslow's sentences. I show that this sentence is provable, using Löb's theorem, as in the standard proof of the Henkin sentence.

Below, I present a few theorems on sentences of arithmetic which are related to Yablo's Paradox approximately as Gödel's first undecidable sentence was related to the Liar paradox. Each one involves taking a sentence—Gödel's, Jeroslow's, or Henkin's, as the superscripts suggest—which is normally thought of as self-referential, and "unfurling" it into a sequence of sentences which would naturally be thought of as constituting a non-well-founded chain of reference.

These results are of interest, first, because they tend to support Gödel's assertion that any "epistemological antinomy" can be used as a guide in finding some form of mathematical incompleteness,[4] and second, because they present a striking link between the behavior of the arithmetical sentences mentioned above (Gödel's, Henkin's, and Jeroslow's) and the behavior of their respective unfurled counterparts considered below. In the absence of a counterexample, the correspondence suggests that the pattern may extend to other cases, or may even be itself a candidate for investigation, should a reasonable formalization of the notion of unfurling become available.

1. PRELIMINARIES

The technology employed by the proofs below is almost entirely standard. We do require free-variable versions of the diagonal lemma, and of Hilbert-Bernays derivability conditions. The generalized derivability conditions (GD1),(GD2), (GD3) may be somewhat unfamiliar. The last two can be found in [Fef60], and the first is easy to prove in any reasonable system. We let M denote any r.e. system of reasonable power (if you like, PA).

Definition 1. We let $\Box$ abbreviate a predicate satisfying

(GD1) $\mathsf{M} \vdash \phi(x) \Rightarrow \mathsf{M} \vdash \Box(\ulcorner \phi(\dot{x}) \urcorner)$
(GD2) $\mathsf{M} \vdash \Box(x \underline{\rightarrow} y) \rightarrow (\Box(x) \rightarrow \Box(y))$
(GD3) $\psi(y_1, \ldots, y_k) \in \Sigma_1 \Rightarrow \mathsf{M} \vdash \psi(y_1, \ldots, y_k) \rightarrow \Box(\ulcorner \psi(\dot{y}_1, \ldots, \dot{y}_k) \urcorner)$

Remark 2 (Generalized Löb's Theorem)**.** It follows from (GD1)-(GD3) that $M \vdash \Box(\ulcorner \phi(\dot{x}) \urcorner) \rightarrow \phi(x)$ implies $\mathsf{M} \vdash \phi(x)$.

[1] Gödel makes this connection in his published presentation of his result—[Göd86, p149].
[2] See [Yab93] for a canonical presentation
[3] presented in [Jer73]
[4] [Göd86, n14]

Definition 3. With k a free variable, let

$$\mathsf{M} \vdash Y^J(k) \leftrightarrow (\forall x > k)[\Box(\ulcorner \neg Y(\dot{x}) \urcorner)]$$
$$\mathsf{M} \vdash Y^G(k) \leftrightarrow (\forall x > k)[\neg \Box(\ulcorner Y(\dot{x}) \urcorner)]$$
$$\mathsf{M} \vdash Y^H(k) \leftrightarrow (\forall x > k)[\Box(\ulcorner Y(\dot{x}) \urcorner)]$$

Remark 4. Inspecting the definitions, we see that if $x > y$, then $\mathsf{M} \vdash Y^{J/G/H}(\bar{y}) \to Y^{J/G/H}(\bar{x})$.

2. Theorems

Theorem 5.

1. *For any* k
   (a) *If* 1-Con(M)*, then* $\mathsf{M} \nvdash Y^J(\bar{k})$.
   (b) *If* Con(M)*, then* $\mathsf{M} \nvdash \neg Y^J(\bar{k})$
2. *For any* k
   (a) *If* Con(M)*, then* $\mathsf{M} \nvdash Y^G(\bar{k})$
   (b) *If* 1-Con(M)*, then* $\mathsf{M} \nvdash \neg Y^G(\bar{k})$
3. $\mathsf{M} \vdash Y^H(k)$

Proof. For (1), first suppose that $\mathsf{M} \vdash Y^J(\bar{k})$. Then, evidently, $\mathsf{M} \vdash \Box(\ulcorner \neg Y^J(\dot{\bar{k}} + 1) \urcorner)$, so by 1-Con$(\mathsf{M})$, $\mathsf{M} \vdash \neg Y^J(\bar{k}+1)$, which contradicts $Y^J(\bar{k})$, by Remark 4. Now, suppose that $\mathsf{M} \vdash \neg Y^J(\bar{k})$. This implies $\mathsf{M} \vdash (\exists x > \bar{k})[\neg \Box(\ulcorner \neg Y(\dot{x}) \urcorner)]$, which violates the second incompleteness theorem.

For (2), first suppose that $\mathsf{M} \vdash Y^G(\bar{k})$. Then, evidently $\mathsf{M} \vdash \neg \Box(\ulcorner Y(\dot{\bar{k}} + 1) \urcorner)$, But, also $\mathsf{M} \vdash Y^G(\bar{k} + 1)$, by Remark 4, and so $\mathsf{M} \vdash \Box(\ulcorner Y^G(\dot{\bar{k}} + 1) \urcorner)$, which violates the consistency of M. On the other hand, suppose $\mathsf{M} \vdash \neg Y^G(\bar{k})$. Then $\mathsf{M} \vdash (\exists x > k)[\Box(\ulcorner Y^G(\dot{x}) \urcorner)]$. By 1-consistency, and Σ_1-completeness we get that, for some x, $\mathsf{M} \vdash \Box(\ulcorner Y^G(\dot{\bar{x}}) \urcorner)$, and by a second application of 1-consistency, we get that $\mathsf{M} \vdash Y^G(\bar{x})$, which is impossible, by the first part of the argument.

For (3), we aim to show that $\mathsf{M} \vdash \Box(\ulcorner Y^H(\dot{k}) \urcorner) \to Y^H(k)$, and appeal to the generalized Löb's theorem of Remark 2. So, assume, in M, that $\Box(\ulcorner Y^H(\dot{k}) \urcorner)$, so that by (GD1),(GD2),

$$\Box(\ulcorner (\forall x > \dot{k})[\Box(\ulcorner Y^H(\dot{x}) \urcorner)] \urcorner). \tag{γ}$$

In M, fix an arbitrary $x > k$. By (GD3), and since "$x > k$" is a Σ_1 formula, we have $\Box(\ulcorner \dot{x} > \dot{k} \urcorner)$. Thus, by ($\gamma$) and (GD1),(GD2), we have $\Box(\ulcorner (\forall z > \dot{x})[\Box(\ulcorner Y^H(\dot{z}) \urcorner)] \urcorner)$. So equivalently, by (GD1), (GD2), $\Box(\ulcorner Y^H(\dot{x}) \urcorner))$. Since $x > k$ was arbitrary, we have $(\forall x > k)[\Box(\ulcorner Y^H(\dot{x}) \urcorner)]$, which implies $Y^H(k)$, Discharging our assumption, we have $\mathsf{M} \vdash \Box(\ulcorner Y^H(\dot{k}) \urcorner) \to Y^H(k)$, whence, by the generalized Löb Theorem 2, $\mathsf{M} \vdash Y^H(k)$. $\Box$

Theorem 6. *Let* k *be a free variable. Then*

$$\mathsf{M} \vdash \mathrm{Con}(\mathsf{M}) \leftrightarrow Y^G(k)$$

Proof. The right-to-left implication is clear. For left-to-right, formalize the argument of (2) above:

That is, in M, let $x > k$ be arbitrary. Then, aiming at a refutation of Con(M), assume in M that $\Box(\ulcorner Y^G(\dot{x}) \urcorner)$. We then have, by (GD1), (GD2), that $\Box(\ulcorner \neg \Box(\ulcorner Y^G(\dot{x} + 1) \urcorner) \urcorner)$, but also $\Box(\ulcorner Y^G(\dot{x} + 1) \urcorner)$, from which, by (GD3), we have $\Box(\ulcorner \Box(\ulcorner Y^G(\dot{x} + 1) \urcorner) \urcorner)$. These two together imply $\neg$ Con(M). So, discharging our assumption and contraposing, Con$(\mathsf{M}) \to \neg \Box(\ulcorner Y^G(\dot{x}) \urcorner)$. As $x > k$ was arbitrary, we have $(\forall x > k)[\mathrm{Con}(\mathsf{M}) \to \neg \Box(\ulcorner Y^G(\dot{x}) \urcorner)]$, so by standard manipulation of quantifiers, Con$(\mathsf{M}) \to (\forall x > k)[\neg \Box(\ulcorner Y^G(\dot{x}) \urcorner)]$. So, Con$(\mathsf{M}) \to Y^G(k)$. $\Box$

Theorem 7. *Let* k *be a free variable. Then,*

$$\mathsf{M} \vdash \mathrm{Con}(\mathsf{M}) \leftrightarrow \neg Y^J(k)$$

Proof. The right-to-left implication is clear. For left-to-right, let $x > k$ be arbitrary in M, and assume in M that $\Box(\ulcorner\neg Y^J(\dot{x})\urcorner)$. Then, by (GD1), (GD2), $\Box(\ulcorner(\exists y > \dot{x})[\neg\Box(\ulcorner\neg Y(\dot{y})\urcorner)]\urcorner)$, whence $\Box\,\mathrm{Con}(\mathsf{M})$. By a formalized version of Gödel's second theorem, $\Box(\ulcorner\mathrm{Con}(\mathsf{M})\urcorner) \to \neg\,\mathrm{Con}(\mathsf{M})$, so $\neg\,\mathrm{Con}(\mathsf{M})$ follows. Thus, discharging our assumption and contraposing, $\mathrm{Con}(\mathsf{M}) \to \neg\Box(\ulcorner Y^J(\dot{x})\urcorner)$. Since $x > k$ was arbitrary, evidently, $\mathrm{Con}(\mathsf{M}) \to (\forall x > k)[\neg\Box(\ulcorner\neg Y^J(\dot{x})\urcorner)]$, so very directly, $\mathrm{Con}(\mathsf{M}) \to (\exists x > k)[\neg\Box(\ulcorner\neg Y^J(\dot{x})\urcorner)]$, whence $\mathrm{Con}(\mathsf{M}) \to \neg Y^J(k)$. $\Box$

3. ACKNOWLEDGEMENTS

Thanks to Rafal Urbaniak and Cezary Cieslinski for some suggested simplifications, particularly for the observation that a standard diagonal lemma, and 1-consistency were sufficient for these results. Thanks to Chris Porter for invaluable editorial advice.

References

[Fef60] Solomon Feferman. Arithmetization of metamathematics in a general setting. *Fundamenta Mathematicae*, 1960. 1

[Göd86] Kurt Gödel. On Formally Undecidable Propositions of Principia Mathematica and Related Systems I. In *Kurt Gödel, Collected Works, Volume I: Publications 1929-1936*. 1986. 1

[Jer73] R G Jeroslow. Redundancies in the Hilbert-Bernays Derivability Conditions for Gödel's Second Incompleteness Theorem. *The Journal of Symbolic Logic*, 1973. 1

[Yab93] Stephen Yablo. Paradox without Self-Reference. *Analysis*, 1993. 1